EFFECT OF VARIOUS IMAGE RECONSTRUCTION PARAMETERS ON PET LESION DETECTION PERFORMANCE

by

Alan Michael Morey

A thesis submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Bioengineering

The University of Utah

May 2015

The University of Utah Graduate School

STATEMENT OF THESIS APPROVAL

The thesis of **Alan Michael Morey**

has been approved by the following supervisory committee members:

Dan J. Kadrmas	, Chair	**01/12/2015** Date Approved
Gregory A. Clark	, Member	**01/08/2015** Date Approved
Christopher R. Butson	, Member	**01/08/2015** Date Approved

and by **Patrick A. Tresco** , Chair/Dean of

the Department/College/School of **Bioengineering**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Positron emission tomography (PET) images can be reconstructed using a wide variety of techniques. Two aspects of image reconstruction are addressed in this thesis: the number of subsets used for the block-iterative ordered-subsets expectation-maximization (OSEM) reconstruction algorithm, and using smaller in-plane pixels. Both of these aspects of PET image reconstruction affect image quality. Although image quality in PET is difficult to quantify, it can be evaluated objectively using task-basked assessments such as lesion detection studies. The objective of this work was to evaluate both the effect of the number of OSEM subsets and pixel size on general oncologic PET lesion detection. Experimental phantom data were taken from the Utah PET Lesion Detection Database Resource, modeling whole-body oncologic ^{18}F-FDG PET imaging of a 92kg patient. The data comprised multiple scans on a Biograph mCT time-of-flight (TOF) scanner, with up to 23 sources modeling lesions (diam. 6-16 mm) distributed throughout the phantom for each scan. Two observer studies were performed as part of this thesis. In the first study, images were reconstructed with maximum-likelihood expectation-maximization (MLEM) and with OSEM using 12 different numbers of subsets (*i.e.*, 2-84 subsets). Localization receiver operating characteristics (LROC) analysis was applied using a mathematical observer. The probability of correct localization (P_{LOC}) and the area under the LROC (A_{LROC}) curve were used as figures-of-merit in order to quantify lesion detection performance. The results demonstrated an

overall decline in lesion detection performance as the number of subsets increased. This loss of image quality can be controlled using a moderate number of subsets (*i.e.*, 12-14 or fewer). In the second study, images were reconstructed with 2.036 mm and 4.073 mm in-plane pixels. Similar LROC analysis methods were applied to determine lesion detection performance for each pixel size. The results of this study demonstrated that images with ~2 mm pixels provided higher lesion detection performance than those with ~4 mm pixels. The primary drawback of using smaller pixels (*i.e.* ~2 mm) was a 4-fold increase in reconstruction time and data storage requirements. Overall, this work demonstrated that reconstructing with moderate subsets or with smaller voxel sizes may provide important benefits for general PET cancer imaging.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to acknowledge my advisor Dan Kadrmas who made it possible for me to accomplish this work. He has spent countless hours teaching me about PET imaging and the methods and qualities required to be an excellent researcher. I would not be where I am today without him. I would also like to acknowledge my committee members, Greg Clark and Chris Butson, who have graciously agreed to give me feedback on this thesis and spend their time teaching me. I also want to acknowledge all the journal reviewers and editors who have made it possible for parts of this work to be published.

I would finally like to thank all my family that has supported me through my endeavors in school. I want to thank my parents who worked very hard to give their kids a better life. I also want to thank my siblings who gave me wonderful examples of hard work and dedication. I want to especially thank my wife Sarah who has been by my side to help me through and through.

CHAPTER 1

INTRODUCTION

Background and Motivation

Positron emission tomography (PET) is a nuclear medicine imaging modality used to produce images of physiological and functional processes in the body. Biologically active molecules are labeled with positron-emitting radioisotopes. These molecules, with attached radioisotopes, are referred to as radiotracers or tracers. These tracers can then be introduced to the body and regional uptake occurs throughout various tissues in the body. Once a positron is emitted, it interacts with a nearby electron to produce a pair of annihilation photons with identical energies (511keV) that are emitted almost antiparallel to each other [1]. Opposing pairs of PET detectors can then detect the pair of photons in coincidence and localize their origin along the line between the two detectors. These coincidence events are then used in image reconstruction to acquire images that show the tracer uptake throughout the body. There is currently a wide range of tracers being researched, while few are being used in clinical practice. The most common PET tracer is ^{18}F-flurodeoxyglucose (FDG), which is used for imaging glucose uptake throughout the body. Since glucose uptake is prevalent in cancerous tissue, PET imaging is often used for the detection, staging, and treatment of cancer.

PET imaging is very complex in nature, and image quality varies widely with

varying imaging systems, reconstruction techniques, and tracers. This can create issues with defining meaningful standards for image quality, which can eventually affect diagnosis, staging, and treatment of cancer clinically. Although image quality is difficult to quantify, image fidelity can be quantified by using image characteristics such as spatial resolution, contrast, and noise. However, these measures of image fidelity may not be predictive of performance in the clinic. Image quality is best defined by how well the image performs for its use in the clinic. Such objective assessment of image quality can be performed by studying the performance of observers for given clinical tasks.

The task relevant to oncologic PET imaging involves detecting a focal warm lesion on a noisy background. This task measures both how well an observer detects actually present lesions (sensitivity), and correctly differentiates noise blobs from lesions that are not present (specificity). Observers have different propensities for detecting lesions. This propensity is referred to as the observer's internal decision threshold. Receiver operating characteristics (ROC) analysis [2], and variants thereof, provide the means for quantifying observer performance across all decision thresholds. This work relies heavily upon localization receiver operating characteristics (LROC) [2-4], which adds upon ROC by taking into account localization of the lesion as well as detection of the lesion. This quantification can provide a meaningful measure of image quality relevant to routine clinical PET oncologic imaging.

Objective and Thesis Outline

Image quality varies widely depending upon imaging system and image reconstruction technique. Two different aspects of image reconstruction were studied as

part of this thesis. The first involves the block-iterative reconstruction algorithm, ordered-subsets expectation-maximization (OSEM) [5-7]. This algorithm involves dividing the projection data into "subsets" that are operated on sequentially during each OSEM iteration. This provides an acceleration factor approximately equivalent to the number of subsets used for image reconstruction. The second aspect of image reconstruction involves decreasing the in-plane pixel size relative to the typical ~4 mm. It has been shown that reconstruction with smaller pixels can improve spatial resolution in the reconstructed images; however, it also affects noise properties [1, 8-9]. The objective of this thesis is to provide an objective evaluation of how these two different reconstruction parameters, specifically OSEM subsets and voxel size, affect lesion detection performance for general oncologic PET imaging with ^{18}F-FDG.

Throughout the course of my thesis research, one paper, entitled, "Effect of Varying Number of OSEM Subsets on PET Lesion Detectability," was published in the *Journal of Nuclear Medicine Technologist*, while another manuscript, entitled, "Effect of Using 2 mm Voxels on Observer Performance for PET Lesion Detection," was submitted to *IEEE Transactions on Nuclear Science* and is currently under review. These papers appear as Chapters 2 and 3, respectively. The methods sections for these chapters are very similar since this work used experimental data from the Utah PET Lesion Detection Database Resource [10]. However, the scientific questions being asked in each work were addressed by testing different aspects of image reconstruction, specifically the number of OSEM subsets and voxel size, in each of these chapters. These scientific questions can also help us to understand whether or not these two reconstruction parameters have any potential to improve health care. The methods for these studies

were designed to allow for comparisons between the reconstruction parameters being studied and the milestone algorithms used in the clinic today. This provides information regarding whether the reconstruction techniques of interest may have any clinical significance. The objective of this work is to answer these scientific questions, which brings knowledge regarding potential ways to improve health care for cancer patients using PET imaging. In the last chapter of this thesis, conclusions are drawn based on the results found in each work, the clinical significance is addressed, and potential avenues of future work are presented.

References

[1] S. R. Cherry, J. A. Sorenson, and M. E. Phelps, *Physics in Nuclear Medicine:* Elsevier Health Sciences, 2012.

[2] L. M. Popescu, "Nonparametric ROC and LROC analysis," *Med Phys*, vol. 34, pp. 1556-64, 2007.

[3] P. Khurd and G. Gindi, "Rapid computation of LROC figures of merit using numerical observers (for SPECT/PET reconstruction)," *IEEE Trans Nucl Sci*, vol. 4, pp. 2516-2520, 2003.

[4] H. C. Gifford, P. E. Kinahan, C. Lartizien, and M. A. King, "Evaluation of multiclass model observers in PET LROC studies," *IEEE Trans Nucl Sci*, vol. 54, pp. 116-123, 2007.

[5] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans Med Imaging*, vol. 13, pp. 601-9, 1994.

[6] B. F. Hutton, H. M. Hudson, and F. J. Beekman, "A clinical perspective of accelerated statistical reconstruction," *Eur J Nucl Med*, vol. 24, pp. 797-808, 1997.

[7] D. J. Kadrmas, "Statistically regulated and adaptive EM reconstruction for emission computed tomography," *IEEE Nucl Sci Symp Conf Rec*, pp. 287-291, 2000.

[8] W. Epley, H. Jacene, M. Lodge, and R. Wahl, "The impact of PET pixel size on contrast recovery for small lesions," *J Nucl Med Meeting Abstracts,* p. 2018, 2009.

[9] G. Shakirin, P. Crespo, F. Fiedler, A. Wagner, and W. Enghardt, "Optimum voxel

size for reconstruction of in-beam PET data," in Conference Record of the 2008 IEEE Nuclear Science Symposium and Medical Imaging Conference, pp. 5066-5069, 2008.

[10] D. J. Kadrmas, "The Utah PET lesion detection database," in Conference Record of the 2009 IEEE Nuclear Science Symposium and Medical Imaging Conference, pp. 3693-3699, 2009.

CHAPTER 2

EFFECT OF VARYING NUMBER OF OSEM SUBSETS ON PET LESION DETECTABILITY

This research was originally published in JNMT. A. M. Morey and D. J. Kadrmas, "Effect of varying number of OSEM subsets on PET lesion detectability," *J Nucl Med Technol,* 2013;41:268-73.

Abstract

Iterative reconstruction has become the standard for routine clinical PET imaging. However, iterative reconstruction is computationally expensive, especially for time-of-flight (TOF) data. Block-iterative algorithms such as ordered-subsets expectation-maximization (OSEM) are commonly used to accelerate the reconstruction. There is a tradeoff between the number of subsets and reconstructed image quality. The objective of this work was to evaluate the effect of varying the number of OSEM subsets on lesion detection for general oncologic PET imaging. Experimental phantom data were taken from the Utah PET Lesion Detection Database, modeling whole-body oncologic ^{18}F-FDG PET imaging of a 92-kg patient. The experiment consisted of 24 scans over 4 d on a TOF PET/CT scanner, with up to 23 lesions (diameter, 6-16 mm) distributed throughout

the thorax, abdomen, and pelvis. Images were reconstructed with maximum-likelihood expectation-maximization (MLEM) and with OSEM using 2-84 subsets. The reconstructions were repeated both with and without TOF. Localization receiver operating characteristic (LROC) analysis was applied using the channelized nonprewhitened observer. The observer was first used to optimize the number of iterations and smoothing filter for each case that maximized lesion detection performance for these data; this was done to ensure that fair comparisons were made with each test case operating near its optimal performance. The probability of correct localization and the area under the LROC curve were then analyzed as functions of the number of subsets to characterize the effect of OSEM on lesion detection performance. Compared with the baseline MLEM algorithm, lesion detection performance with OSEM declined as the number of subsets increased. The decline was moderate out to about 12-14 subsets and then became progressively steeper as the number of subsets increased. Comparing TOF with non-TOF results, the magnitude of the performance drop was larger for TOF reconstructions. PET lesion detection performance is degraded when OSEM is used with a large number of subsets. This loss of image quality can be controlled using a moderate number of subsets (*e.g.*, 12-14 or fewer), retaining a large degree of acceleration while maintaining high image quality. The use of more aggressive subsetting can result in image quality degradations that offset the benefits of using TOF or longer scan times.

Introduction

Iterative reconstruction algorithms that model Poisson statistics have become the standard for routine clinical PET imaging. Maximum-likelihood expectation-

maximization (MLEM) is the foundational algorithm; however, it is computationally expensive and requires many iterations to reach a suitable image. This problem is exacerbated by the emergence of time-of-flight (TOF) imaging, where the computational cost per iteration can be an order of magnitude slower than non-TOF [1]. Block-iterative algorithms such as ordered-subsets expectation-maximization (OSEM) are widely used to accelerate iterative image reconstruction [2-8]. Here, the projection data are divided into subsets that are operated on sequentially during each OSEM iteration. The number of subsets provides the approximate acceleration factor—one iteration of OSEM with N subsets provides an image roughly similar to that from N iterations of MLEM [4-6]. However, there is a tradeoff between the number of subsets and image quality. When the number of subsets is large, the size of each subset is small and each contains less tomographic and statistical information, potentially resulting in enhanced noise structures and other subset-related artifacts in the final image [4].

When OSEM is used in the clinic, it is important to understand the tradeoff between increasing the number of subsets (more acceleration) and image quality degradations (noise, artifacts). One approach would be to study how spatial resolution, contrast, and noise are affected by changing the number of subsets; however, these measures of image fidelity do not necessarily predict performance for clinical tasks. The accepted approach for objectively evaluating image quality in PET is to perform task-based assessments in which the different images are evaluated in terms of an observer's ability to perform a given task, such as detecting a lesion in the image. This task includes both detecting a lesion that is actually present (sensitivity) and correctly ruling out noise blobs that are not lesions (specificity) [9]. The objective of this work was to evaluate the

relationship between the number of OSEM subsets and image quality in terms of lesion-detectability for general oncologic PET imaging with ^{18}F-FDG.

Our group has established techniques for evaluating PET lesion detection performance using specially designed phantom experiments [10-13], and these data and methodologies have been combined in a resource called the Utah PET Lesion Detection Database Resource [14]. The resource consists of experimental data and routines for performing localization receiver operating characteristic (LROC) analysis [15-17] with the channelized nonprewhitened (CNPW) numeric observer [18]. Model observers such as the CNPW have been shown to correlate with human observers for simple lesion detection tasks [10, 11, 18-24], and they offer the ability to quickly and repeatedly review large numbers of images. These data and LROC methods have previously been used to evaluate PET lesion detection performance when modeling the point spread function [10], using TOF data [11], and reducing scan times [13].

In this work, experimental data from the Utah PET Lesion Detection Database were reconstructed with the MLEM algorithm (*i.e.*, 1 subset) as a baseline and with OSEM using 11 different numbers of subsets (2-84 subsets). The reconstructions were repeated both with and without TOF data. Lesion detection performance was assessed for each case using the CNPW observer with LROC analysis. The following sections describe the experimental data, reconstruction and data processing, LROC study methods, and results. The effect of increasing the number of OSEM subsets on lesion-detectability is then analyzed, and conclusions based on the results are drawn.

Materials and Methods

Experimental Data for Lesion Detection Assessment

The lesion-detectability study used experimental data from the Utah PET Lesion Detection Database [14] for the custom large whole-body phantom scanned on a Biograph mCT TOF PET/CT scanner (Siemens Medical Solutions) with timing resolution 527.5 ± 4.9 ps [25]. The phantom, shown in Figure 2.1, had 3 main components: a 3-dimensional brain phantom; an anthropomorphic thorax phantom containing liver, lungs, and rib cage; and a pelvis with bladder compartment. The approximate dimensions of the phantom are 43 × 28.0 cm at the widest points, and total length is approximately 83.1 cm. Accounting for the missing mass of the arms and legs, the phantom models an approximately 92 kg patient. The phantom also had several custom modifications designed to increase realism for modeling whole-body oncologic ^{18}F-FDG PET [14].

The experiment consisted of 6 back-to-back whole-body scans acquired each day over the 4 d of the experiment. Each whole-body scan acquired list-mode data for 4 min. per bed position over 6 bed positions. Three of the 4 d had 21-23 shell-less ^{68}Ge (half-life, 270.8 d) sources modeling lesions [26] with diameters of 6-16 mm distributed throughout the phantom lungs, liver, and soft-tissue compartments (mediastinum, abdomen, pelvis) to model tumors with focal ^{18}F-FDG uptake. On the final day, no lesions were present, providing true-negative images for the observer study. This multiscan protocol provided numerous images and lesions with varying count levels and lesion target-to-background ratios. The overall activity levels for the 6 scans broadly covered the full range of activity levels representative of sites administering 370-555

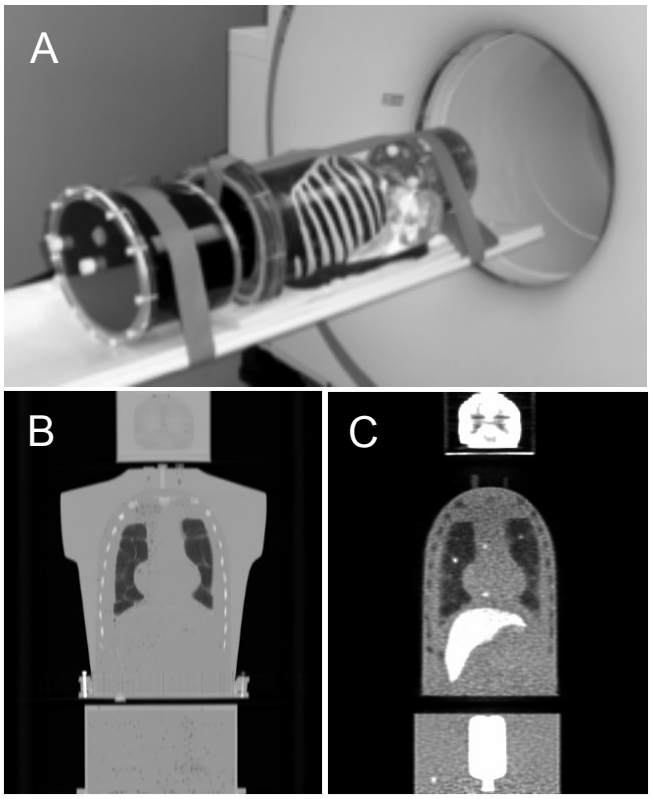

Fig. 2.1 The whole-body phantom, shown on the PET/CT scanner table (A), consists of a brain compartment; thorax with liver, lungs, and rib cage/spine; and pelvis with bladder. It models a patient of approximately 92 Kg. Coronal CT (B) and PET (C) images show the main phantom compartments and structures. Example lesions can also be seen in the PET image in both lungs, the mediastinum, and pelvis regions.

MBq ^{18}F-FDG with uptake times ranging from 60 to 120 min.

Image Reconstruction and Processing

The raw scan data, including list-mode files, attenuation maps, scanner calibrations, and scatter and randoms estimates, were loaded to an offline workstation and reconstructed using manufacturer-supplied software (Siemens Medical Solutions). The baseline reconstruction algorithm was ordinary Poisson line-of-response MLEM with spatially variant point spread function modeling [27], and each scan was reconstructed both with and without TOF. The reconstructed image matrix was 168 × 168, with 4.073 mm pixels and 2.027 mm slice thickness. After reconstructing with MLEM, the reconstructions were repeated using OSEM with every available number of subsets. The sinogram data had 168 angles, and the reconstruction software required that the number of angles per subset be a multiple of 2, giving the following numbers of subsets: 2, 3, 4, 6, 7, 12, 14, 21, 28, 42, and 84. Thus, 12 non-TOF and 12 TOF reconstructions were performed for each scan: MLEM and 11 versions of OSEM covering 2-84 subsets.

One challenge in comparing different OSEM reconstructions is that the rate of iterative convergence depends on the number of subsets, and similarly the noise properties (and hence the best postreconstruction filter) also depend on the number of subsets and iterations. To provide a fair comparison, it was important to objectively select the number of iterations and filter used for each case. The standard approach used with the Utah PET Lesion Detection Database [10-14] is to empirically optimize the number of iterations and postreconstruction filter for each algorithm that maximizes lesion detection performance for that algorithm. As such, each algorithm was run out to

at least 120 MLEM-equivalent iterations (*e.g.*, 20 iterations for OSEM6), with a minimum of 10 iterations for each case, and the intermediate images from each iteration were stored for subsequent processing and analysis. The optimal number of iterations and smoothing filter were then selected using preliminary LROC studies.

The true location of each lesion in the phantom was determined from phantom setup coordinate grids and was confirmed on the CT scans. As reported previously [13], scans 2-5 were found to provide the most clinically representative activity and noise levels, and data from these scans were used for the remainder of the study. This provided a total of 268 lesion-present test images (21-23 lesions × 4 scans/d × 3 d with lesions present) plus 268 corresponding lesion-absent test images (from the scans without lesions) to be used for the LROC study for each reconstruction algorithm.

LROC Analysis

Preliminary LROC studies were first used to select the optimal number of iterations and postreconstruction filter for each algorithm, ensuring that each algorithm was fairly compared at near-maximum performance. Here, 21 different 3-dimensional Gaussian filters were applied to the images from each iteration, with SD ranging from 0.0 (no filter) to 2.0 voxels in 0.1-voxel increments. The area under the LROC curve (A_{LROC}) was computed for each iteration–filter combination. Figure 2.2 shows how A_{LROC} changed as a function of iteration and filter for 2 TOF reconstruction cases and demonstrates that local changes to iteration and filter have minimal effects on A_{LROC}. The iteration and filter that maximized A_{LROC} were identified and selected for each algorithm; these values are listed in Table 2.1. These parameters maximized A_{LROC} for

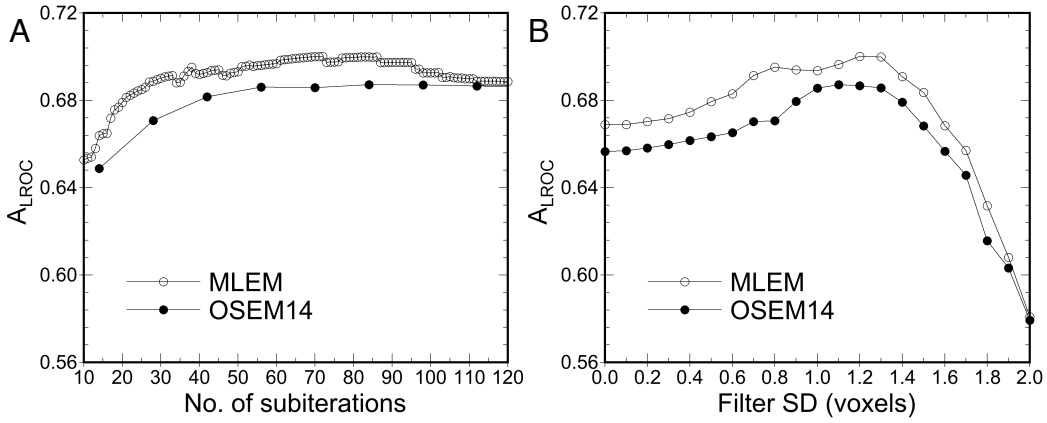

Fig. 2.2 Example analysis results used for selecting the number of iterations and filter strength for each case studied. The plot on the left (A) shows A_{LROC} *vs.* subiteration for MLEM and OSEM14 (where 1 subiteration represents 1 full pass through the data; *i.e.*, 1 iteration MLEM = 1 subiteration, and 1 iteration OSEM14 = 14 subiterations). Here the data are shown for the filter that maximized A_{LROC} at each subiteration. The analogous plot on the right (B) shows A_{LROC} *vs.* filter SD, where each datum is shown for the number of iterations that maximized A_{LROC} for that filter strength. These data represent a portion of the multidimensional sampling used to optimize the number of iterations and filter strength for the phantom data used in this work.

Table 2.1
Selected Reconstruction Parameters

	Non-TOF		**TOF**	
No. Subsets	No. Iterations	Filter SD	No. Iterations	Filter SD
1	120	1.3	72	1.2
2	58	1.3	36	1.2
3	37	1.4	23	1.2
4	24	1.2	21	1.2
6	19	1.4	7	0.9
7	18	1.4	13	1.3
12	10	1.3	5	1.2
14	10	1.5	6	1.1
21	7	1.2	3	0.9
28	7	1.4	3	1.2
42	4	1.2	2	0.8
84	5	1.3	1	0.8

this particular set of experimental data, and they do not necessarily represent near-optimal or optimal parameters for clinical use. The topic of optimizing the number of iterations and filters for clinical use is large and complex and falls outside the scope of this work.

Empiric selection of the best number of iterations and filters required reading 7,834,176 test images to cover 268 lesion-present and lesion-absent test cases for each algorithm, iteration, and filter. It would not have been feasible to read this many images with human observers; however, the CNPW numeric observer completed this task within a few days. The CNPW observer computes a numeric rating, analogous to a human observer's confidence level, regarding the presence or absence of a lesion at each image location. The location with the highest rating was selected as the most-probable lesion location for the LROC analysis. Additional details on the CNPW observer [18, 21] and its training and application to our experimental phantom data [10-13] can be found in the references. As in this prior work, a radius of correct localization equal to 2.5 voxels was found to correctly identify hits while minimizing random localizations and was used throughout this study. Two figures of merit were used to quantify lesion detection performance: the probability of correct localization (P_{LOC}) and A_{LROC}. P_{LOC} is simply the fraction of lesions correctly localized within the 2.5-voxel threshold. A_{LROC} plots the correctly localized true-positive fraction versus the false-positive fraction, computed from the observer rating data and known truth. Higher values for these measures indicate higher lesion detection performance.

Results

Example Images

Example images reconstructed for each number of OSEM subsets are shown in Figure 2.3. The MLEM image provides the baseline for comparison and corresponds to OSEM with 1 subset. Increasing the number of subsets resulted in increased noise and subtle shape artifacts in these images, especially for the highest numbers of subsets. The overall objective of this work was to evaluate how these changes in the images affect lesion detection performance for general oncologic PET imaging. Consider, for example, the sample images shown in Figure 2.4. This case had a true 8 mm lesion in the left lung and noise blobs of similar size and contrast in the mediastinum. The use of OSEM with 28 subsets resulted in lower contrast for the true (lung) lesion as compared with MLEM, coupled with increased contrast of the mediastinal noise blob. In this example, the CNPW observer correctly identified the lung lesion (true-positive) on the MLEM image but falsely identified the mediastinal noise blob (false-positive) on the OSEM28 image. This example illustrates how subset-related artifacts can affect lesion detection performance.

Lesion Detectability *vs.* Number of Subsets

Figure 2.5 presents the main results of this paper, showing how P_{LOC} and A_{LROC} changed as functions of the number of OSEM subsets. Lesion detection performance declined overall as the number of subsets increased. The decline was moderate out to about 12-14 subsets and then became progressively steeper as the number of subsets increased. When TOF and non-TOF results were compared, the same trend in

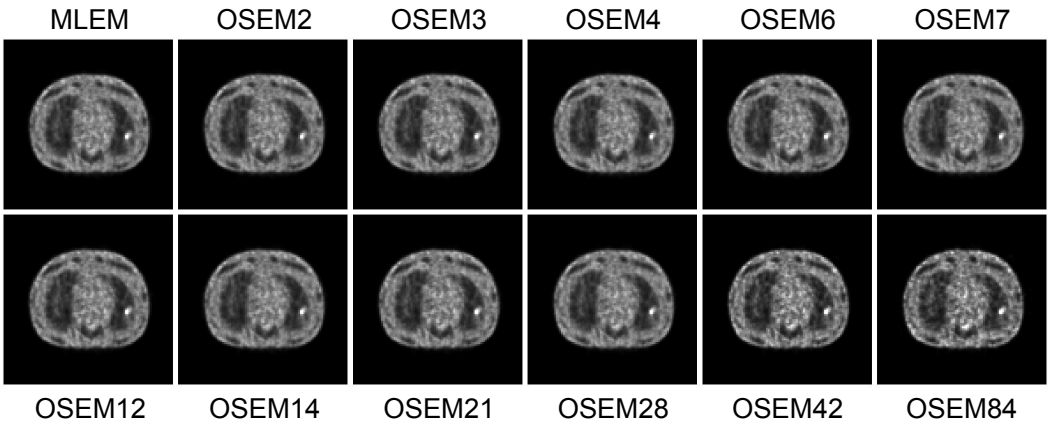

Fig. 2.3 Example reconstructed images with TOF for each number of OSEM subsets, showing a slice in the mediastinum with a 10 mm diameter hot lesion in the left lung. Each image is shown at approximately 56 MLEM-equivalent iterations. Increasing noise and subtle shape-related artifacts can be observed in the images as the number of subsets increases.

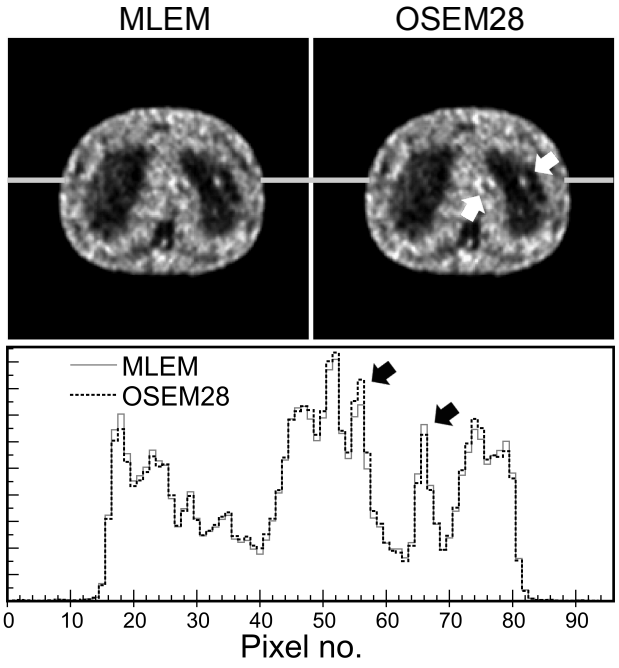

Fig. 2.4 Example TOF images for MLEM and OSEM28 with optimal iteration and filter as determined by this study, demonstrating potential effects on lesion-detectability. The focus in the left lung is a true 8 mm hot lesion, and the foci in the mediastinum are noise artifacts. Horizontal profiles showing relative intensity (arbitrary units) show that OSEM28 resulted in a loss of contrast for the lung lesion (right black arrow), coupled with an increase in contrast for the mediastinal noise blobs (left black arrow), as compared to MLEM. This example represents a case where the observer identified the correct (lung) lesion on the MLEM image (*i.e.*, true-positive reading), but misidentified the mediastinal noise blob as a lesion on the OSEM28 image (*i.e.*, false-positive).

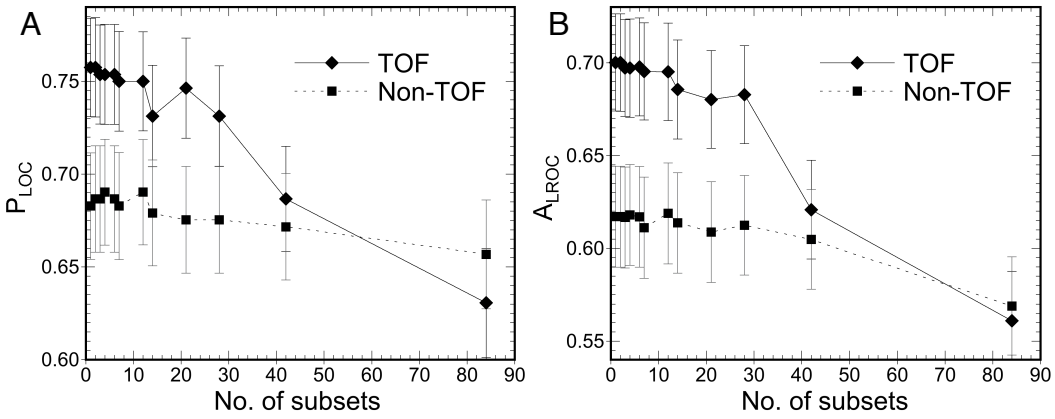

Fig. 2.5 Lesion detection performance, as measured by P_{LOC} (A) and A_{LROC} (B), plotted as a function of the number of OSEM subsets for both TOF and non-TOF reconstructions. Performance declined overall as the number of subsets increased, with a marked drop in performance beyond approximately 28 subsets. For the TOF reconstructions, the performance drop at 42 subsets effectively canceled the benefit of TOF.

performance was observed, but the magnitude of the performance drop was much larger for TOF. Overall, these results demonstrate that lesion detection performance is only slightly degraded when a moderate number of subsets is used, suggesting that acceleration factors of as much as approximately 10 times can be safely attained with OSEM. However, more aggressive subsetting can cause more significant losses in image quality and adversely affect lesion-detectability.

Discussion

When LROC studies are performed, it is important to provide a context for interpreting the magnitude of differences in the figures of merit (*i.e.*, in P_{LOC} and A_{LROC}) in clinically relevant terms. The absolute magnitudes of P_{LOC} and A_{LROC} are determined largely by the experimental design. For example, one could include many large, high-contrast lesions that are easily detected—pushing the values of P_{LOC} and A_{LROC} close to one for all algorithms studied. Conversely, one could include many small, low-contrast lesions in the test dataset, resulting in P_{LOC} and A_{LROC} values closer to zero. Ideally, the test dataset would exactly model the clinically encountered distribution, in which case the absolute magnitude of the results would impart clinical meaning; however, such a distribution is not well understood and would vary widely by disease state. Furthermore, such a distribution would include many always-detectable lesions (found by all test algorithms) as well as many invisible lesions (*e.g.*, micrometastases), neither of which would add to the statistical power of the study for differentiating the test algorithms. The lesion test data used here, as for most lesion-detectability studies, were designed to provide high statistical power for differentiating and ranking the test algorithms studied.

As such, the differences in the results should be interpreted within a meaningful context.

We provide 2 such contexts in this work. First, the impact of TOF versus non-TOF on PET lesion detection performance has previously been evaluated in both phantoms and patients and is becoming well understood [11, 13, 28-30]. Comparing the TOF versus non-TOF results in Figure 2.5, one sees that the TOF reconstruction with 42 subsets provided approximately the same lesion detection performance as the non-TOF reconstruction with MLEM. In essence, the degradation of using 42 subsets balanced and offset the benefit of using TOF in these data. The degradation from using 28 subsets cost approximately 20% of the benefit of TOF. Although these results should not be construed as exact quantifications, they do provide a context for assessing the significance of the changes observed in the results.

To provide additional context for interpreting the results, we repeated the MLEM TOF reconstructions and computed A_{LROC} as a function of scan time. Here, the raw list-mode PET data files were statistically pruned from 240 s per bed position to 180, 120, and 90 s per bed position (corresponding to whole-body scan times of 24, 18, 12, and 9 min, respectively). The technique was the same as that presented in a previous publication [13]. Repeating the LROC analysis for these images, we computed the change in A_{LROC} as a function of scan time for MLEM. The results are shown in Figure 2.6, plotted alongside the results for changing the number of subsets. Here, using OSEM with 21 subsets was found to result in the same loss of detectability as found for MLEM when the scan time was reduced from 240 to approximately 205 s per bed position. Overall, these data suggest that reconstructing with OSEM up to about 12-14 subsets has only a moderate effect on lesion detection performance but that using more subsets can

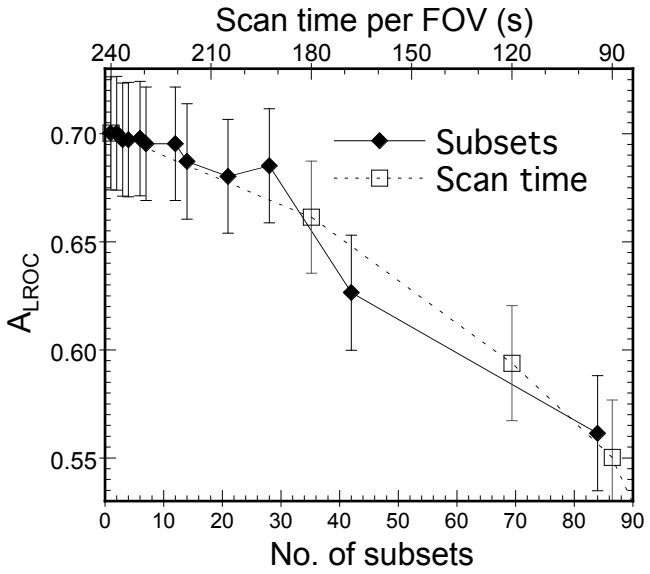

Fig. 2.6 Comparison of how lesion detection performance, as quantified by A_{LROC}, is affected by increasing the number of OSEM subsets or decreasing the scan time for TOF reconstructions. These data provide a context for interpreting the significance of the changes in A_{LROC} observed in this work. For example, the use of approx. 32 subsets would result in the same loss of performance as shortening the scan time from 240 to 180 sec. per bed position.

result in more significant degradations.

Conclusion

When OSEM is used for tomographic reconstruction, the number of subsets provides the approximate acceleration factor for this algorithm as compared with MLEM. However, increasing the number of subsets also results in increased noise and subset-related artifacts in the image. This work evaluated the effect of changing the number of OSEM subsets on lesion detection performance for general oncologic PET imaging. As compared with the baseline MLEM algorithm, lesion detection performance declined as the number of OSEM subsets increased. The decline was moderate out to approximately 12-14 subsets for the data studied here, beyond which performance dropped more rapidly with the number of subsets. TOF PET reconstructions showed greater effect than non-TOF reconstructions. The degree of loss of lesion-detectability with 21 subsets was similar to that observed when the scan time was reduced from 240 to 205 s per bed position. Similarly, the use of 42 subsets with TOF data offset the value of TOF, resulting in the same A_{LROC} as non-TOF reconstructed with MLEM. We conclude that PET lesion detection performance is degraded when OSEM is used with a large number of subsets for both non-TOF and TOF reconstructions. This loss of image quality can be controlled by using a moderate number of subsets (*e.g.*, 12-14 or fewer), retaining a large degree of acceleration while maintaining high image quality.

References

[1] G. Pratx, S. Surti, and C. Levin, "Fast list-mode reconstruction for time-of-flight PET using graphics hardware," *IEEE Trans Nucl Sci*, vol. 58, pp. 105-109, 2011.

[2] R. Boellaard, A. van Lingen, and A. A. Lammertsma, "Experimental and clinical evaluation of iterative reconstruction (OSEM) in dynamic PET: quantitative characteristics and effects on kinetic modeling," *J Nucl Med*, vol. 42, pp. 808-17, 2001.

[3] J. Dey and M. A. King, "Theoretical and numerical study of MLEM and OSEM reconstruction algorithms for motion correction in emission tomography," *IEEE Trans Nucl Sci*, vol. 56, pp. 2739-2749, 2009.

[4] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans Med Imaging*, vol. 13, pp. 601-9, 1994.

[5] B. F. Hutton, H. M. Hudson, and F. J. Beekman, "A clinical perspective of accelerated statistical reconstruction," *Eur J Nucl Med*, vol. 24, pp. 797-808, 1997.

[6] D. J. Kadrmas, "Statistically regulated and adaptive EM reconstruction for emission computed tomography," *IEEE Nucl Sci Symp Conf Rec*, pp. 287-291, 2000.

[7] D. J. Kadrmas, "LOR-OSEM: statistical PET reconstruction from raw line-of-response histograms," *Phys Med Biol*, vol. 49, pp. 4731-44, 2004.

[8] X. Liu, C. Comtat, C. Michel, P. Kinahan, M. Defrise, and D. Townsend, "Comparison of 3-D reconstruction with 3D-OSEM and with FORE+OSEM for PET," *IEEE Trans Med Imaging*, vol. 20, pp. 804-14, 2001.

[9] H. H. Barrett and K. J. Myers, Foundations of Image Science. Hoboken, New Jersey: John Wiley & Sons, Inc., 2004.

[10] D. J. Kadrmas, M. E. Casey, N. F. Black, J. J. Hamill, V. Y. Panin, and M. Conti, "Experimental comparison of lesion detectability for four fully-3D PET reconstruction schemes," *IEEE Trans Med Imaging*, vol. 28, pp. 523-34, 2009.

[11] D. J. Kadrmas, M. E. Casey, M. Conti, B. W. Jakoby, C. Lois, and D. W. Townsend, "Impact of time-of-flight on PET tumor detection," *J Nucl Med*, vol. 50, pp. 1315-23, 2009.

[12] D. J. Kadrmas and P. E. Christian, "Comparative evaluation of lesion detectability for 6 PET imaging platforms using a highly reproducible whole-body phantom with (22)Na lesions and localization ROC analysis," *J Nucl Med*, vol. 43, pp. 1545-54, 2002.

[13] D. J. Kadrmas, M. B. Oktay, M. E. Casey, and J. J. Hamill, "Effect of scan time on oncologic lesion detection in whole-body PET," *IEEE Trans Nucl Sci*, vol. 59, pp. 1940-1947, 2012.

[14] P. Khurd and G. Gindi, "Fast LROC analysis of Bayesian reconstructed emission tomographic images using model observers," *Phys Med Biol*, vol. 50, pp. 1519-32, 2005.

[15] L. M. Popescu, "Nonparametric ROC and LROC analysis," *Med Phys*, vol. 34, pp. 1556-64, 2007.

[16] P. Khurd and G. Gindi, "Rapid computation of LROC figures of merit using numerical observers (for SPECT/PET reconstruction)," *IEEE Trans Nucl Sci*, vol. 4, pp. 2516-2520, 2003.

[17] H. C. Gifford, P. E. Kinahan, C. Lartizien, and M. A. King, "Evaluation of multiclass model observers in PET LROC studies," *IEEE Trans Nucl Sci*, vol. 54, pp. 116-123, 2007.

[18] K. J. Myers and H. H. Barrett, "Addition of a channel mechanism to the ideal-observer model," *J Opt Soc Am A*, vol. 4, pp. 2447-57, 1987.

[19] H. C. Gifford, M. A. King, D. J. de Vries, and E. J. Soares, "Channelized hotelling and human observer correlation for lesion detection in hepatic SPECT imaging," *J Nucl Med*, vol. 41, pp. 514-21, 2000.

[20] H. C. Gifford, M. A. King, P. H. Pretorius, and R. G. Wells, "A comparison of human and model observers in multislice LROC studies," *IEEE Trans Med Imaging*, vol. 24, pp. 160-9, 2005.

[21] R. D. Fiete, H. H. Barrett, W. E. Smith, and K. J. Myers, "Hotelling trace criterion and its correlation with human-observer performance," *J Opt Soc Am A*, vol. 4, pp. 945-53, 1987.

[22] H. H. Barrett, J. Yao, J. P. Rolland, and K. J. Myers, "Model observers for assessment of image quality," *Proc Natl Acad Sci U S A*, vol. 90, pp. 9758-65, 1993.

[23] C. Lartizien, P. E. Kinahan, and C. Comtat, "Volumetric model and human observer comparisons of tumor detection for whole-body positron emission tomography," *Acad Radiol*, vol. 11, pp. 637-48, 2004.

[24] B. W. Jakoby, Y. Bercier, M. Conti, M. E. Casey, B. Bendriem, and D. W. Townsend, "Physical and clinical performance of the mCT time-of-flight PET/CT scanner," *Phys Med Biol*, vol. 56, pp. 2375-89, 2011.

[25] J. J. Hamill, C. E. Arnsdorff, M. E. Casey, L. Xinli, and W. J. Raulston, "A Ge-68 PET hot-sphere phantom with no cold shells," *IEEE Nucl Sci Symp Conf Rec*, pp. 1609-1613, 2005.

[26] V. Y. Panin, F. Kehren, C. Michel, and M. Casey, "Fully 3-D PET reconstruction with system matrix derived from point source measurements," *IEEE Trans Med Imaging*, vol. 25, pp. 907-21, 2006.

[27] J. S. Karp, S. Surti, M. E. Daube-Witherspoon, and G. Muehllehner, "Benefit of time-of-flight in PET: experimental and clinical results," *J Nucl Med*, vol. 49, pp.

462-70, 2008.

[28] S. Surti, J. Scheuermann, G. El Fakhri, M. E. Daube-Witherspoon, R. Lim, N. Abi-Hatem, *et al.*, "Impact of time-of-flight PET on whole-body oncologic studies: a human observer lesion detection and localization study," *J Nucl Med*, vol. 52, pp. 712-9, 2011.

[29] G. El Fakhri, S. Surti, C. M. Trott, J. Scheuermann, and J. S. Karp, "Improvement in lesion detection with whole-body oncologic time-of-flight PET," *J Nucl Med*, vol. 52, pp. 347-53, 2011.

CHAPTER 3

EFFECT OF USING 2 MM VOXELS ON OBSERVER PERFORMANCE FOR PET LESION DETECTION

A. M. Morey, F. Noo, and D. J. Kadrmas, submitted to IEEE Transactions on Nuclear Science (TNS) and currently under review.

Abstract

Positron emission tomography (PET) images are typically reconstructed with an in-plane pixel size of approximately 4 mm for cancer imaging. The objective of this work was to evaluate the effect of using smaller pixels on general oncologic lesion detection. A series of observer studies was performed using experimental phantom data from the Utah PET Lesion Detection Database, which modeled whole-body FDG PET cancer imaging of a 92kg patient. The data comprised 24 scans over 4 days on a Biograph mCT time-of-flight (TOF) PET/CT scanner, with up to 23 lesions (diam. 6-16 mm) distributed throughout the phantom each day. Images were reconstructed with 2.036 mm and 4.073 mm pixels using ordered-subsets expectation-maximization (OSEM) both with and without point spread function (PSF) modeling and TOF. Detection performance was assessed using the channelized nonprewhitened numerical observer with localization receiver operating characteristic (LROC) analysis. Tumor

localization performance and the area under the LROC curve were then analyzed as functions of the pixel size. In all cases, the images with ~2 mm pixels provided higher detection performance than those with ~4 mm pixels. The degree of improvement from the smaller pixels was larger than that offered by PSF modeling for these data, and provided roughly half the benefit of using TOF. These results were confirmed by two human observers, who read subsets of the test data. This study suggests that a significant improvement in tumor detection performance for PET can be attained by using smaller voxel sizes than commonly used at many centers. The primary drawback is a 4-fold increase in reconstruction time and data storage requirements.

Introduction

Positron emission tomography (PET) images are typically reconstructed with an in-plane pixel size of ~4 mm for many general oncologic imaging applications, and ~2 mm for brain imaging. Reconstruction with smaller pixels has been found to improve spatial resolution and contrast recovery in reconstructed PET images; however, it also affects image noise properties [1-3] and is more computationally expensive. Advances in reconstruction algorithms, computer processing speeds, and storage media have made routine reconstruction with smaller pixel sizes feasible for routine use in the clinical setting.

Although image spatial resolution, contrast, and noise are affected by changing the voxel size, these measures of image fidelity are not necessarily predictive of performance for clinical tasks such as detection and staging of cancer. Image quality can be objectively evaluated using task-based assessments which quantify an observer's

ability to perform a task such as detecting a focal, hot lesion on a structured noisy background. This task, relevant to general oncologic PET imaging, includes both determining whether a lesion is actually present (sensitivity) and correctly ruling out noise blobs that are not lesions (specificity) [4]. Receiver operating characteristics (ROC) analysis, and variants thereof, can be used to quantify the observer's performance for lesion-detectability tasks in meaningful measures.

Our group has developed techniques for evaluating general oncologic lesion-detectability in PET using whole-body phantom experiments [5-8], and these methodologies and data have been combined into the Utah PET Lesion Detection Database Resource [9]. The resource contains experimental data useful for performing localization receiver operating characteristics (LROC) studies [10-12] with both the channelized nonprewhitened (CNPW) mathematical observer [13] and human observers. These data and LROC methods have been previously used to evaluate the effects of modeling the point spread function (PSF) [5], TOF [6], and varying the number of OSEM subsets used for iterative reconstruction [8].

The objective of this work was to evaluate the effect of reconstructing with smaller voxel sizes (*e.g.*, ~2 mm in-plane, as compared to ~4 mm) on lesion detection performance for general oncologic PET imaging. Lesion detection performance was assessed using standardized metrics, and the effect of decreasing pixel size was evaluated for images reconstructed both with and without PSF modeling and TOF. The following sections describe the experimental data, reconstruction and processing techniques, LROC study methods, and results. The effect of using smaller pixels on lesion detection is then analyzed, and conclusions are drawn based on these data.

Methods

Experimental Phantom Data

The lesion detection study used experimental data from the Utah PET Lesion Detection Database Resource [9] for a custom large whole-body phantom scanned on a Biograph mCT TOF PET/CT scanner (Siemens Medical Solutions) with timing resolution 527.5 ± 4.9 ps [14]. The phantom, shown in Fig. 3.1, has three main components: a 3-dimensional (3D) brain phantom; anthropomorphic thorax phantom containing liver, lungs, and rib cage; and a pelvis with bladder compartment. The approximate dimensions of the phantom are 43.0 × 28.0 cm at the widest points, and the total length is approximately 83.1 cm. Accounting for the missing mass of the arms and legs, this phantom models a patient of approximately 92 Kg. The phantom also has a number of custom modifications designed to increase realism for modeling whole-body general oncologic imaging with ^{18}F-fluorodeoxyglucose (FDG).

The experimental data consisted of 4 days of experiment, with six back-to-back whole-body scans acquired each day. The overall activity levels for the six scans broadly covered the full range of activity levels representative of sites administering 3.4–10.6 mCi FDG with uptake times ranging from 60 to 120 min. Each scan acquired listmode data for 4 min. per bed over six bed positions. Three of the four days had 21-23 ‘‘shell-less’’ ^{68}Ge ($T_{1/2}$ = 270.8d) sources [15], diameters 6-16 mm, distributed throughout the phantom lungs, liver, and soft tissue compartments (mediastinum, abdomen, pelvis). These sources modeled tumors with focal FDG uptake with tumor:background ratios ranging from 1.9 to 5.9 in the various phantom compartments and scans. The scans performed on the final day had no lesions present, providing true-negative images for the

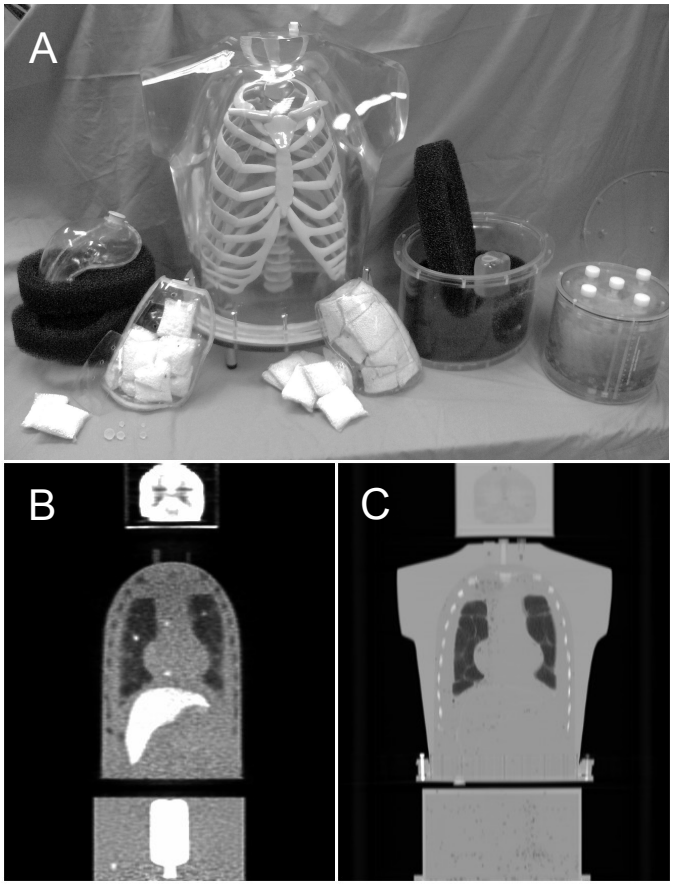

Fig. 3.1 The whole-body phantom (A) consists of a brain compartment; thorax with liver, lungs, and rib cage/spine; and pelvis with bladder. It models a patient of approximately 92 Kg. Coronal PET (B) and CT (C) images show the main phantom compartments and structures. Example lesions can also be seen in the PET image in the lungs, the mediastinum, and pelvis regions.

observer studies. This multiscan protocol provided numerous images with varying count levels and lesion contrasts, as demonstrated by maximum intensity projection images shown in Fig. 3.2. Lesion locations and activities were designed to cluster near the verge of detectability, maximizing statistical power for differentiating the test algorithms in the observer studies.

Image Reconstruction

The raw scan data, including listmode files, attenuation maps, scanner calibrations, and scatter and randoms estimates, were reconstructed offline using manufacturer-provided reconstruction software. Images were reconstructed with line-of-response (LOR) ordered-subsets expectation-maximization (OSEM) with 14 subsets out to 12 iterations, both with and without spatially variant PSF modeling [16] and TOF. The algorithms analyzed throughout this study are referred to as LOR-OSEM (baseline), PSF, TOF, and PSF+TOF. Corrections for scanner normalization, deadtime, attenuation, scatter, and randoms were applied using the manufacturer-provided reconstruction software.

The reconstructions with each algorithm were repeated with two in-plane pixel sizes: 4.073 mm and 2.036 mm, referred to as "4 mm" and "2 mm" throughout this paper. The corresponding reconstructed image matrix sizes were 168 × 168 and 336 × 336, respectively. In all cases, the slice thickness was 2.027 mm. The images for all algorithms and both pixel sizes, including the intermediate iterations, were stored for subsequent processing and analysis.

The true location of each lesion in the phantom was determined from phantom

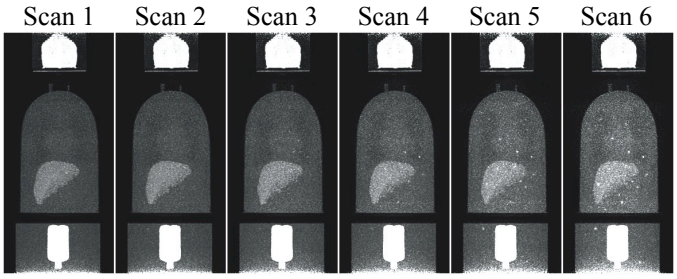

Fig. 3.2 Example maximum intensity projection (MIP) images with lesions present. The upper limit of the grayscale was lowered to enhance visualization of the body compartments. The images show increasing lesion contrast for each successive scan as the ^{18}F background decayed while the ^{68}Ge lesions remained ~constant. This provided multiple contrasts for each lesion, ranging from undetectable to borderline to easily visible, increasing statistical power for the observer studies.

setup coordinate grids and confirmed on the CT scans. A total of 402 lesion-present test cases (21–23 lesions × 6 scans/day × 3 days with lesions present) were used for each algorithm and pixel size, along with the corresponding 402 lesion-absent test image slices taken from the scans acquired without lesions.

LROC Studies with CNPW Observer

The CNPW observer, as developed by Gifford *et al.* [13, 17], was used with the LROC study to compute a perception rating and most-likely lesion location for each test image. Additional details regarding the CNPW observer and its training and application to our experimental phantom data can be found in [5-8, 13, 17]. Two versions of the CNPW observer were used in this work: a "2D" observer which read single slices of the image (centered at lesion-center), and a "3D" observer which read 7 contiguous image slices (also centered at lesion-center). The two observers were identical in all other respects.

Figure 3.3 shows the probability of correct lesion localization (P_{LOC}) plotted as a function of the localization radius acceptance threshold. A radius threshold value of 10.182 mm was found to correctly identify 'hits' while minimizing random localizations, and this threshold was used throughout the study. Two figures-of-merit were used for quantifying lesion detection performance: P_{LOC} and the area under the LROC curve (A_{LROC}). Here, P_{LOC} is the fraction of lesions correctly localized within the 10.182 mm threshold, or more simply the fraction of lesions found by the observer. A_{LROC} is the area under the LROC curve, which plots the correctly-localized true-positive fraction *vs.* the false-positive fraction, and is computed from the observer rating data and known truth.

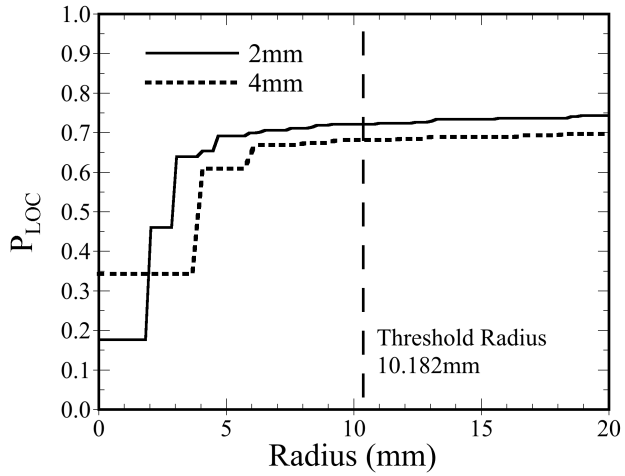

Fig. 3.3 The fraction of lesions correctly localized by the CNPW observer, plotted for each pixel size as a function of radius of correct localization for the PSF+TOF algorithm. A radius threshold of 10.182 mm was used in this work to determine correct localization.

Higher values for these metrics indicate higher lesion detection performance.

Preliminary LROC studies using the CNPW mathematical observer were performed in order to select near-optimal parameters. There were 21 postreconstruction 3-dimensional (3D) Gaussian filters applied to the images for each iteration, with filter width (standard deviation, SD) ranging from 0.0 (no filter) to 8.15 mm in 0.41 mm increments. The A_{LROC} was computed for each iteration-filter combination. Since changing the pixel size affects both the rate of iterative convergence and noise properties, the iteration number and postreconstruction filter were optimized for all cases in order to ensure that each test case was evaluated at near-optimal performance. Figure 3.4 shows example data for TOF images demonstrating how A_{LROC} changed as a function of iteration and filter for both pixel sizes. In order to ensure that the images for both pixel sizes were being evaluated with near-optimal processing parameters, the iteration number and 3D Gaussian filter strength combination that maximized A_{LROC} for each algorithm and pixel size was identified and used for the LROC study with the CNPW observer. This empirical optimization of the number of iterations and filter strengths required reading 1,620,864 test images to cover 402 lesion-present and lesion-absent test cases for each algorithm, iteration and filter. The resultant parameter values used for the CNPW observer study are listed in Table 3.1.

Effect of Changing Pixel Size

The CNPW observer results for the four algorithms and optimal reconstruction parameters were compared across both pixel sizes in order to determine the effect of reducing the pixel size upon P_{LOC} and A_{LROC} for each of the four algorithms. The

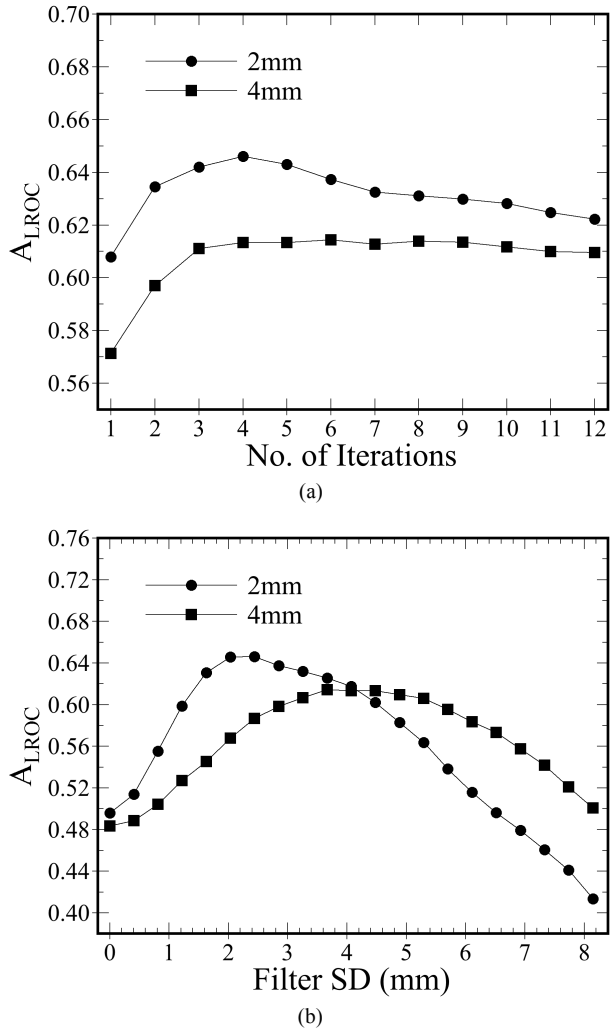

Fig. 3.4 Example analysis TOF results used for selecting number of iterations and filter strength. Top plot (a) shows A_{LROC} *vs.* iteration for both pixels studied. Here, the data are shown for filters that maximized A_{LROC} at each iteration. Analogous plot below (b) shows A_{LROC} *vs.* filter SD, where each datum is shown for number of iterations that maximized A_{LROC} for that filter strength. These data represent a portion of multidimensional sampling used to optimize the number of iterations and filter strength for phantom data used in this work.

Table 3.1
Selected Reconstruction Parameters

	4 mm		2 mm	
Model	No. of iterations	Filter SD (mm)	No. of iterations	Filter SD (mm)
2D CNPW Observer Study				
LOR-OSEM	12	5.29	7	2.85
PSF	12	5.29	12	4.07
TOF	6	3.67	4	2.44
PSF+TOF	6	4.07	9	2.85
3D CNPW Observer Study				
LOR-OSEM	11	2.44	6	2.44
PSF	11	2.44	12	3.67
TOF	5	0.41	6	1.63
PSF+TOF	12	2.04	9	2.04
Human Observer Study				
TOF	4	3.67	4	4.07
PSF+TOF	5	3.26	5	3.67

uncertainty in each figure-of-merit was estimated as the standard deviation over 10,000 bootstrap estimates. The paired-sample Tukey HSD multiple comparison test was then used to test the null hypotheses that the A_{LROC} for 2 mm and 4 mm pixels were the same, versus the alternative hypothesis that 2 mm pixels had higher A_{LROC} than 4 mm pixels, for each of the four reconstruction algorithms. These tests were performed with significance level $\alpha = 0.05$. Note that statistical tests comparing the different reconstruction algorithms (LOR *vs.* PSF *vs.* TOF) were not performed, as these algorithms have previously been compared and the objective of the current work is to evaluate the effect of using 2 mm *vs.* 4 mm pixels.

Key results from the CNPW observer were then confirmed using human observers (as in [17-20]), where manageable-sized subsets of the test data for the TOF and PSF+TOF algorithms at both pixel sizes were read by 2 human observers. Notably, the number of iterations and filter strengths used for the human observer studies, shown in Table 3.1, were heuristically selected to be representative of clinical practices at our institution, rather than the rather high 'optimal' number of iterations identified by the CNPW observer (see Table 3.1 for comparison).

The CNPW results were first used to identify subgroups of the test images to be read by the human observers that provided challenging detection tasks across all reconstruction algorithms studied. As in previous work [7], scans 2-5 of the 6 scans acquired each day of experiment were found to provide count levels most representative of clinical scans. Of the 268 lesion-present test cases in scans 2-5, the CNPW observer was used to exclude lesions that were either always missed or always found, resulting in a set of 200 test images (100 lesion-present plus 100 corresponding lesion-absent test

cases). This process was intended to maximize statistical power of the human observer study by maximizing the number of informative test cases without including such a large number of test cases that observer fatigue became significant. These images were randomly divided in 40 training images and 160 test images for the TOF and PSF+TOF algorithms at each pixel size. Two experienced medical physicists acted as observers for the studies. Note that these observers were not trained PET clinicians; however, previous work demonstrated that such observers are appropriate for the lesion detection task studied herein [6, 21].

The human observers were blinded to which algorithm was presented, and both the ordering of the test cases and images presented were randomized. The observers performed two tasks on each image. First, the location determined to be the most likely to contain a lesion was selected by a mouse click. Second, a confidence rating was selected on a 6-point scale ranging from 1 (high confidence lesion absent) to 6 (high confidence lesion present). The observers were informed that approximately half of the test images would contain lesions, and that each image would have exactly 1 or 0 lesions present, but there could be many noise blobs present. For each test case, the observers first underwent a training session by reading 40 training images. Here, the observers were immediately provided with the truth regarding lesion presence and location after reading each image. Each training session was immediately followed by the test session for the same algorithm and pixel size, where no feedback was provided after reading each image.

The LROC curves for each observer were computed using the nonparametric approach of Popescu [12] with Epanechnikov kernel. The fraction of lesions found

(P_{LOC}) and area under the LROC curve (A_{LROC}) for each observer were also computed for each observer, and then averaged to obtain the final results. The methods used were the same as those previously developed and used for the Utah PET Lesion Detection Database Resource, and additional details can be found in [5, 6]. As with the CNPW observer results, a paired-sample Tukey HSD multiple-comparison test was performed to determine statistically significant differences in A_{LROC} for the different pixel sizes.

Results

Figure 3.5 shows example reconstructed images at both pixel sizes for the TOF and PSF+TOF algorithms. Each image contains one lesion, and no filter was applied to these images so that the differences in noise texture can be visually assessed. Visual differences in spatial resolution, contrast, and background noise can be observed in the images. These differences display different image characteristics for the two pixel sizes, in particular differences in noise texture and lesion contrast.

Figure 3.6 provides a more detailed example of images with an 8 mm lesion in the left lung, reconstructed with the parameters determined for this study (Table 3.1). The contrast of the lesion was markedly higher for the image with 2 mm pixels than for the image with 4 mm pixels; however, the background noise was also somewhat higher for the smaller pixels. The LROC studies performed in this work objectively assess how these differences in image characteristics affect observer performance for the lesion detection task.

The results of the study are presented in Table 3.2, with key results highlighted in Fig. 3.7. Lesion detection performance for all reconstruction algorithms improved when

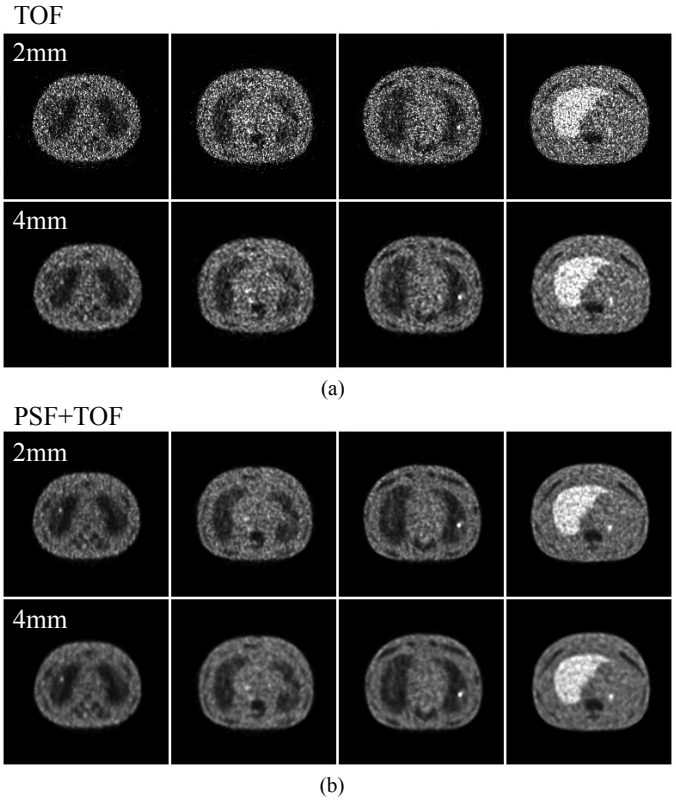

Fig. 3.5 Example unfiltered reconstructed images with 2 mm and 4 mm pixels for both TOF (a) and PSF+TOF (b) reconstructions. Each image contains exactly one lesion (from left-to-right: right lung, mediastinum, left lung, and abdomen). Visual differences in noise textures and spatial resolution properties are evident for the two pixel sizes.

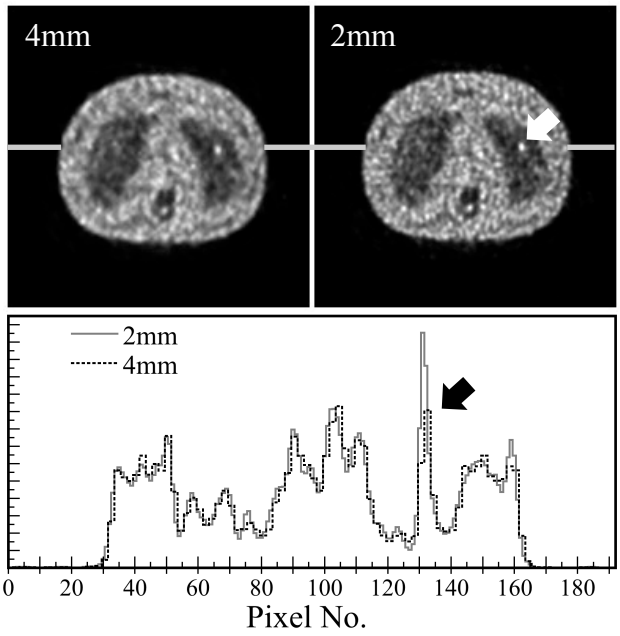

Fig. 3.6 Example reconstructed images with 4 mm and 2 mm pixels and optimum iteration and filter as determined by this study, demonstrating potential effects on lesion detection. Focus in the left lung (white arrow) is a true 8 mm hot lesion. Horizontal profiles showing relative intensity (arbitrary units) demonstrate that using 2 mm pixels resulted in an increase in lesion contrast (black arrow), as compared to 4 mm pixels.

Table 3.2
CNPW Observer Results

Model	P_{LOC} ± SD		A_{LROC} ± SD		Tukey HSD Test (P)[a]
	4 mm	2 mm	4 mm	2 mm	
2D CNPW Observer					
LOR-OSEM	0.629 ± 0.024	0.669 ± 0.024	0.566 ± 0.027	0.592 ± 0.023	0.001
PSF	0.639 ± 0.024	0.667 ± 0.024	0.572 ± 0.023	0.598 ± 0.023	0.005
TOF	0.677 ± 0.023	0.711 ± 0.023	0.614 ± 0.023	0.646 ± 0.022	<0.001
PSF+TOF	0.684 ± 0.023	0.724 ± 0.022	0.628 ± 0.023	0.663 ± 0.022	<0.001
3D CNPW Observer					
LOR-OSEM	0.649 ± 0.024	0.679 ± 0.023	0.574 ± 0.023	0.592 ± 0.023	0.041
PSF	0.654 ± 0.024	0.669 ± 0.023	0.581 ± 0.023	0.596 ± 0.023	0.072 (NS[b])
TOF	0.692 ± 0.023	0.724 ± 0.022	0.626 ± 0.023	0.656 ± 0.022	0.002
PSF+TOF	0.706 ± 0.023	0.739 ± 0.022	0.645 ± 0.023	0.666 ± 0.022	0.008

[a]Tukey HSD multiple comparisons test performed on A_{LROC} figure-of-merit
[b]NS – not significant

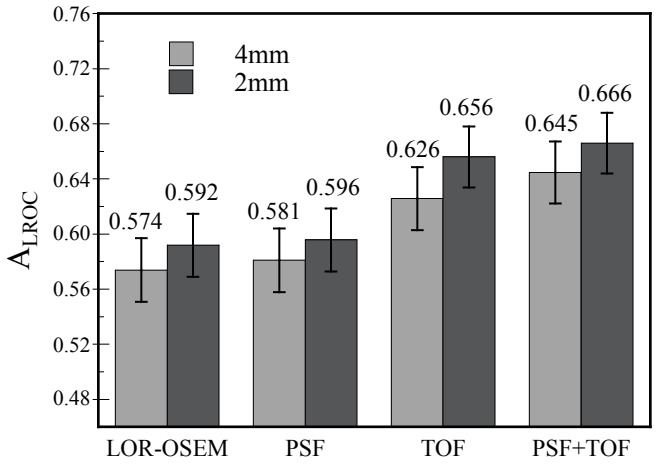

Fig. 3.7 Results for the 3D CNPW observer showing lesion detection performance, as measured by A_{LROC}, shown for 2 mm and 4 mm pixels and each of the reconstruction algorithms studied. Lesion detection performance was significantly higher using 2 mm pixel size for all cases studied.

using 2 mm pixels as compared to 4 mm pixels for both the 2D and 3D CNPW observers for all algorithms. The difference was statistically significant ($p<0.05$) for all cases except the PSF algorithm with 3D CNPW observer ($p=0.07$). The magnitude of the differences in A_{LROC} for 2 mm *vs.* 4 mm pixels was somewhat lower for the 3D CNPW observer as compared to the 2D CNPW observer, but remained statistically significant for 3 of the 4 algorithms studied. Since improvement using smaller pixels was measured for all algorithms, these results suggest that the use of smaller pixels brings value regardless of whether or not PSF modeling and/or TOF are used. The magnitude and significance of these improvements are discussed in more detail in the Discussion section.

The main results from the CNPW mathematical observer study were further evaluated with two human observers, who read manageable-sized subsets of the test data as described in the Methods section. These results are shown in Table 3.3 and Fig. 3.8. The human observer results were consistent with the numerical observer results, finding statistically significant improvement in A_{LROC} when using 2 mm *vs.* 4 mm pixels for both algorithms studied. Note that the absolute value of the human and CNPW results should not be compared with each other, as the human observers read only a subset of the images that the CNPW observer did. However, the results show the same trend and similar magnitude of improvement when using 2 mm as compared to 4 mm voxels, demonstrating consistent results comparing these pixel sizes.

Discussion

When LROC studies are performed, it is important to provide a context for interpreting the magnitude of differences in the figures-of-merit (*i.e.*, in P_{LOC} and A_{LROC})

Table 3.3
Human Observer Validation Study Results

Model	P_{LOC} ± SD		A_{LROC} ± SD		Tukey HSD Test (P)[a]
	4mm	2mm	4mm	2mm	
Human Observer Average					
TOF	0.750 ± 0.035	0.838 ± 0.029	0.656 ± 0.034	0.734 ± 0.031	0.022
PSF+TOF	0.757 ± 0.034	0.832 ± 0.029	0.662 ± 0.034	0.739 ± 0.030	0.015

[a]Tukey HSD multiple comparisons test performed on A_{LROC} figure-of-merit for human observer average only

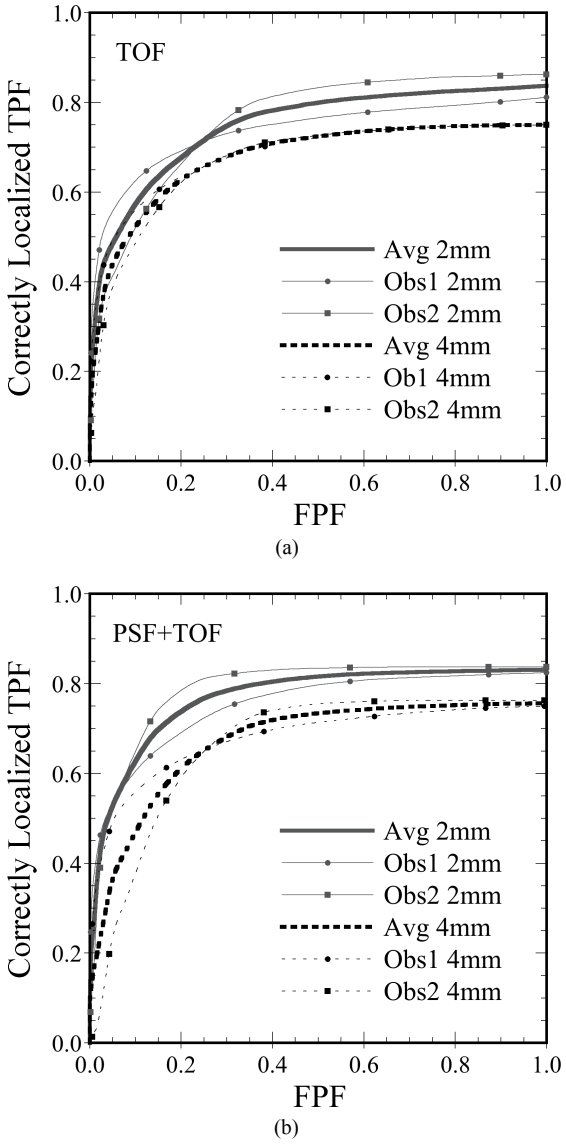

Fig. 3.8 LROC curves for the human observer study for 2 mm and 4 mm pixels for the TOF (a) and PSF+TOF (b) algorithms. The plots show the correctly localized true-positive fraction (TPF) as a function of the false-positive fraction (FPF). These human observer results demonstrate improved lesion detection performance when using 2 mm pixels, confirming the results for the CNPW mathematical observer.

in clinically relevant terms. Comparison of the different algorithms provides a context for evaluating the degree of improvement attained by using smaller pixels. Reconstructing with 2 mm pixels provided a greater degree of improvement in these data than that provided by PSF modeling, and it provided approximately half the degree of improvement as measured for TOF. Notably, the degree of improvement offered by using the smaller pixel size was similar regardless of whether or not PSF modeling was used, and all three reconstruction differences (pixel size, PSF, and TOF) provided cumulative improvements. This suggests that reconstructing with smaller pixels, PSF modeling, and TOF all utilize fundamentally different mechanisms for improving image quality for lesion-detectability.

To provide an additional context for interpreting the results, we repeated the PSF+TOF reconstructions and CNPW LROC studies as a function of scan time. Here, the raw list-mode PET data files were statistically pruned from 240 seconds per bed position to 180, 120, and 90 seconds per bed position (corresponding to whole-body scan times of 24, 18, 12, and 9 min., respectively). This technique for contextualizing the magnitude of LROC study results has been previously established [7]. By repeating the LROC analysis for these images, we computed the change in A_{LROC} as a function of scan time for both 2 mm and 4 mm pixels. The results are shown in Fig. 3.9. These data demonstrate that decreasing the pixel size from 4 mm to 2 mm provides a degree of improvement in lesion detection performance similar to increasing the scan time by approximately 25-33% per bed position.

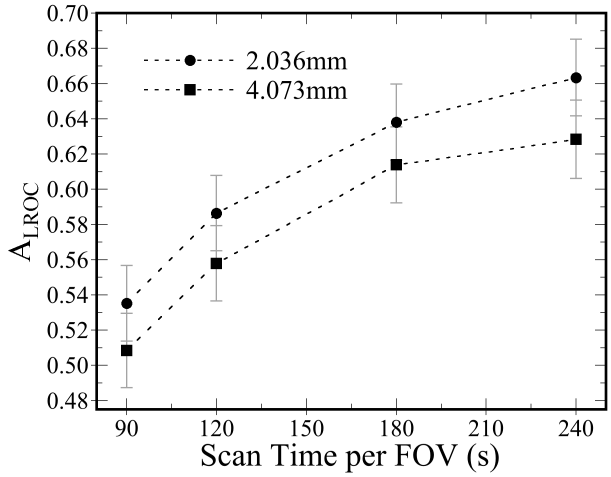

Fig. 3.9 Comparison differences in lesion detection performance for changing pixel sizes versus changing scan time for PSF+TOF reconstructions. These data provide a context for interpreting the significance of changes in A_{LROC} observed in this work. For example, use of 2 mm pixels instead of 4 mm pixels provided an improvement in performance roughly equivalent to increasing the scan time 25-33% per bed position.

Summary and Conclusion

This study evaluated how the use of smaller pixels affects lesion detection performance in general oncologic PET imaging. The results demonstrate that reconstructing with smaller pixel sizes (*i.e.*, ~2 mm instead of ~4 mm) can significantly improve detection performance for focal lesions in a noisy background. The degree of improvement observed here was greater than that offered by PSF modeling, and was approximately half of that offered by TOF; however, the relative magnitude of these differences depends in part on the phantom size used and may differ in broader situations. Improved performance when using smaller pixels was observed regardless of whether PSF modeling or TOF was used, suggesting that each utilizes different mechanisms to improve detection performance. The degree in improvement by using 2 mm pixels was also similar to that observed by increasing the scan time approximately 25-33% per bed position. The primary drawbacks of using 2 mm instead of 4 mm pixels were approximately 4-fold increases in reconstruction time and image storage requirements. These results demonstrate that reconstructing with smaller voxel sizes may provide an important benefit for general PET cancer imaging applications.

References

[1] S. R. Cherry, J. A. Sorenson, and M. E. Phelps, *Physics in Nuclear Medicine:* Elsevier Health Sciences, 2012.

[2] W. Epley, H. Jacene, M. Lodge, and R. Wahl, "The impact of PET pixel size on contrast recovery for small lesions," *J Nucl Med Meeting Abstracts,* p. 2018, 2009.

[3] G. Shakirin, P. Crespo, F. Fiedler, A. Wagner, and W. Enghardt, "Optimum voxel size for reconstruction of in-beam PET data," in Conference Record of the 2008 IEEE Nuclear Science Symposium and Medical Imaging Conference, pp. 5066-5069, 2008.

[4] H. H. Barrett and K. J. Myers, *Foundations of Image Science.* Hoboken, New Jersey: John Wiley & Sons, Inc., 2004.

[5] D. J. Kadrmas, M. E. Casey, N. F. Black, J. J. Hamill, V. Y. Panin, and M. Conti, "Experimental comparison of lesion detectability for four fully-3D PET reconstruction schemes," *IEEE Trans Med Imag,* vol. 28, pp. 523-34, 2009.

[6] D. J. Kadrmas, M. E. Casey, M. Conti, B. W. Jakoby, C. Lois, and D. W. Townsend, "Impact of time-of-flight on PET tumor detection," *J Nucl Med,* vol. 50, pp. 1315-23, 2009.

[7] D. J. Kadrmas, M. B. Oktay, M. E. Casey, and J. J. Hamill, "Effect of scan time on oncologic lesion detection in whole-body PET," *IEEE Trans Nucl Sci,* vol. 59, pp. 1940-1947, 2012.

[8] A. M. Morey and D. J. Kadrmas, "Effect of varying number of OSEM subsets on PET lesion detectability," *J Nucl Med Technol,* vol. 41, pp. 268-73, 2013.

[9] D. J. Kadrmas, "The Utah PET lesion detection database," in Conference Record of the 2009 IEEE Nuclear Science Symposium and Medical Imaging Conference, pp. 3693-3699, 2009.

[10] P. Khurd and G. Gindi, "Rapid computation of LROC figures of merit using numerical observers (for SPECT/PET reconstruction)," *IEEE Trans Nucl Sci,* vol. 4, pp. 2516-2520, 2003.

[11] P. Khurd and G. Gindi, "Fast LROC analysis of Bayesian reconstructed emission tomographic images using model observers," *Phys Med Biol,* vol. 50, pp. 1519-32, 2005.

[12] L. M. Popescu, "Nonparametric ROC and LROC analysis," *Med Phys,* vol. 34, pp. 1556-64, 2007.

[13] H. C. Gifford, P. E. Kinahan, C. Lartizien, and M. A. King, "Evaluation of multiclass model observers in PET LROC studies," *IEEE Trans Nucl Sci,* vol. 54, pp. 116-123, 2007.

[14] B. W. Jakoby, Y. Bercier, M. Conti, M. E. Casey, B. Bendriem, and D. W. Townsend, "Physical and clinical performance of the mCT time-of- flight PET/CT scanner," *Phys Med Biol,* vol. 56, pp. 2375-89, 2011.

[15] J. J. Hamill, C. E. Arnsdorff, M. E. Casey, L. Xinli, and W. J. Raulston, "A Ge-68 PET hot-sphere phantom with no cold shells," in Conference Record of the 2005 IEEE Nuclear Science Symposium and Medical Imaging Conference, pp. 1609-1613, 2005.

[16] V. Y. Panin, F. Kehren, C. Michel, and M. Casey, "Fully 3-D PET reconstruction with system matrix derived from point source measurements," *IEEE Trans Med*

Imaging, vol. 25, pp. 907-21, 2006.

[17] H. C. Gifford, M. A. King, P. H. Pretorius, and R. G. Wells, "A comparison of human and model observers in multislice LROC studies," *IEEE Trans Med Imag,* vol. 24, pp. 160-9, 2005.

[18] R. D. Fiete, H. H. Barrett, W. E. Smith, and K. J. Myers, "Hotelling trace criterion and its correlation with human-observer performance," *J Opt Soc Am A,* vol. 4, pp. 945-53, 1987.

[19] H. C. Gifford, M. A. King, D. J. deVries, and E. J. Soares, "Channelized hotelling and human observer correlation for lesion detection in hepatic SPECT imaging," *J Nucl Med,* vol. 41, pp. 514-21, 2000.

[20] C. Lartizien, P. E. Kinahan, and C. Comtat, "Volumetric model and human observer comparisons of tumor detection for whole-body positron emission tomography," *Acad Radiol,* vol. 11, pp. 637-48, 2004.

[21] D. J. Kadrmas and P. E. Christian, "Comparative evaluation of lesion detectability for 6 PET imaging platforms using a highly reproducible whole-body phantom with (22)Na lesions and localization ROC analysis," *J Nucl Med,* vol. 43, pp. 1545-54, 2002.

CHAPTER 4

CONCLUSIONS

The purpose of this thesis was to evaluate how various aspects of image reconstruction can affect image quality for general oncologic PET imaging with ^{18}F-FDG. Evaluating image quality can be difficult to quantify meaningfully. Many researchers rely upon spatial resolution, contrast, and noise to determine a metric of image quality; however, these measures of image fidelity may not be predictive of clinically relevant tasks. This thesis relies heavily on using task-based assessment with LROC analysis to quantify an observer's performance for lesion detection tasks. These metrics provide meaningful measures of image quality for clinically relevant tasks.

The two works included in this thesis evaluated how the number of OSEM subsets and the pixel size affect lesion detection performance in general oncologic PET imaging. The results demonstrate that increasing the number of OSEM subsets led to increased noise and subset-related artifacts in the images. This resulted in decreased lesion detection performance with aggressive subsetting (greater than 21 subsets). The results also demonstrate that reconstructing with smaller pixel sizes (*i.e.*, ~2 mm instead of ~4 mm) can significantly improve lesion detection performance. The primary drawbacks to using smaller pixels include increases in reconstruction time and image storage requirements; however, advances in computer processing speeds and storage

media have made routine reconstruction with smaller pixels feasible. Overall, reconstructing with moderate subsets (*i.e.*, 12-14) or with smaller voxel sizes may provide important benefits for general PET cancer imaging.

In this work, we were able to meaningfully quantify image quality using various image reconstruction techniques; however, it is also important to understand how and if the results found in this work have any clinical significance. In other words, we want to know if the image reconstruction techniques we are studying have the potential to improve health care, specifically diagnostic imaging for cancer patients. Clinical significance is difficult to establish since a large body of research and knowledge is required. When the research is established and commonly accepted, changes can then be made in the clinic if they have the potential to improve patient care. This work adds to that large body of research and provides a way to compare these image reconstruction techniques to techniques that currently have clinical significance.

The experimental design of this work allowed us to test whether or not the image reconstruction techniques studied, specifically OSEM subsets and voxel size, have potential clinical significance. For example, in the voxel size study (Chapter 3), we found improvements in lesion-detectability when decreasing the voxel size. The differences were statistically significant ($p<0.05$) for all cases except the PSF algorithm with 3D CNPW observer ($p=0.07$). Although we determined statistical significance, clinical significance was still unclear. In order to determine whether the differences were potentially clinically significant, we compared the magnitude of the improvements for smaller voxel size with the magnitude of improvement observed using the milestone reconstruction techniques of the past decade, namely point spread function (PSF)

modeling and time-of-flight (TOF).

Since PSF and TOF are commonly used in the clinic and widely accepted, we wanted to see how image quality improvements attained from using them compared to using the reconstruction techniques studied in this work. For example, we determined that using smaller voxel sizes has potential clinical significance since it provided approximately double the improvement of PSF and approximately half the improvement of TOF. Also, we have seen that two of the major PET scanner vendors have implemented smaller voxels sizes for image reconstruction. I have high hopes that this work can and will improve health care for cancer patients. This work demonstrates that the reconstruction techniques studied have the potential to improve lesion-detectability, which can lead to improvements in detection, diagnosis, and staging of cancer.

There are many potential avenues of future research. This work could be extended to include many other aspects of image reconstructions, but only a few are discussed here. Since only two in-plane pixel sizes were evaluated in this work, it would be of interest to evaluate lesion detection performance with more pixel sizes and perhaps varying slice thickness. Throughout the course of this thesis research, various sensitivities to filter strength were observed depending upon the reconstruction algorithm. Future work could also include evaluating the effect of filter and filter strength on lesion detection performance. It would also be valuable to repeat this evaluation of lesion detection performance using patient data to verify that the trends in lesion detection performance are similar.